1970 V (153)

Psychological Review
1970, Vol. 77, No. 6, 585–590

AN OBJECTIVE APPROACH TO SUBJECTIVE EXPERIENCE:

FURTHER EXPLANATION OF A HYPOTHESIS[1]

R. W. SPERRY[2]

Division of Biology, California Institute of Technology

An attempt is made to clarify a hypothesis of consciousness in which the phenomena of subjective experience are conceived to exert a direct causal influence on brain activity. The present article responds to some questions posed by Bindra that arose from a recent statement by Sperry of the concept. The aim is to clarify where possible with material that supplements rather than repeats previous accounts. The hypothesis is compared with other existing theories, and reference is made to a more simplified explanation for nonspecialists presented five years ago. Hopefully the three articles together will reinforce each other to provide a reasonably comprehensible coverage of the interpretation as seen to date.

The puzzlement expressed by Bindra (1970) over my recent statement on consciousness appears to arise in no small part from my not having reiterated therein some of the earlier thinking contained in previous papers (Sperry, 1952, 1965, 1966b) to which only a passing reference was included. In regard to the first main point raised by Bindra, for example, in which he explains that the interpretation of consciousness as an emergent phenomenon is not new, it was made quite clear in the earlier presentations (Sperry, 1965, 1966b) that there was no thought of implying novelty on this score with respect either to vitalism or to consciousness. I explained that this aspect of the present hypothesis has been accepted by many authorities and mentioned Morgan and Herrick as historical examples among proponents of this approach. In describing consciousness in the recent 1969 statement as something "different from and more than" its neural components, I used the quotation marks expressly because this phrase is such an old and familiar one that it seemed reasonable to assume most readers would recognize it as such, including its historical connotations. It did not occur to me that readers of the *Psychological Review* might infer any claim to novelty in this regard. When it is further implied by Bindra, in this same connection, that the interpretation of consciousness as an emergent property is not only not new but also not *controversial*, the criticism goes too far. Those who heretofore have described consciousness as an emergent phenomenon have done so in tentative and speculative terms and have been in a minority. It has been more popular in behavioral science and philosophy both to conceive of consciousness as an "inner aspect" of the brain process or as being "identical" with the neural process, or in terms of psychoneural parallelism as some kind of "epiphenomenon" or "paraphenomenon," or simply to write it off as a "semantic pseudoproblem." The interpretation as an emergent phenomenon has been and still remains today definitely controversial. In the past, the hypothesized emergent phenomena of consciousness have generally been conceived to be passive, parallelistic, and epiphenomenal. It is both new and controversial in our present scheme to suggest that subjective experiences in the form of emergent phenomena actively govern the flow pattern of neural excitation.

Another point of confusion raised by Bindra concerns the possible implication in my hypothesis of dualism, the unprofitable explanatory disadvantages of which are then expounded. Again, the earlier discussion (Sperry, 1966b) was quite explicit on this point stating that the proposed scheme "would eliminate the old dualistic confusions, dichotomies, and paradoxes, proposing instead *a single unified system* [italics added] extending from subnuclear forces near the bottom up through ideas at the top. It would provide a long-sought unifying view on which to base our conception of human nature [p. 5]." That the proposed scheme was not dualistic I hoped would be evident also from its description in terms of the neural events involved in the perception of pain. Pain as a subjective

[1] Preparation of this paper was supported by grants from the Stone Foundation and from the National Institute of Mental Health (Grant MH-03372), and by the F. P. Hixon Fund of the California Institute of Technology. The author thanks Jerre Levy, Colwyn Trevarthen, and Oliver Zangwill for helpful comments on the manuscript.

[2] Requests for reprints should be sent to R. W. Sperry, Division of Biology, California Institute of Technology, Pasadena, California 91109.

experience was explained as a holistic property of a particular spatiotemporal pattern of cerebral excitation that as a dynamic functional entity directly determines the further course of brain activity. The same point was made again (Sperry, 1965, 1966b) when outlining the concept further with reference to the different kinds of control in the hierarchy of command that exists in brain function extending upward through successive organizational levels from the atomic, molecular, neuronal, and nerve circuit levels to that of cerebral processes. Consciousness was described as being a part of this same material hierarchy consisting of dynamic properties that emerge in its upper levels.

Exclusion of dualism in any traditional sense was intended in the later statement as well by describing consciousness as a dynamic property of cerebral excitation that is inseparably tied to, and a direct property of, the brain process, and not a disembodied or supernatural agent. In particular, the conscious phenomena were stated explicitly (Sperry, 1969) to exist within the brain, not as mere epiphenomena but as direct properties of the brain process. A primary object of the proposed scheme is to give consciousness a functional and causal role in brain processes without becoming involved in the dualistic difficulties of the older mentalistic interpretations.

More generally, my statement is held by Bindra to be confusing because it "entangles" the idea of consciousness as an emergent phenomenon with the idea that consciousness plays a causal role in determining neural activity and that these two propositions are independent, neither implying the other. It would be more accurate to say that the hypothesis "combines" these two ideas, and it does so with deliberate and calculated intent. It is true, of course, that the two ideas do not logically imply each other. If they did, there would be nothing new involved. To miss this is to miss a central aspect of the argument.

The proposed interpretation would satisfy the common intuitive impression that conscious phenomena are both real and operationally useful. It would serve also to bring mind as a distinct phenomenon back into the domain of objective science—unlike most other scientific treatments of consciousness that have prevailed since the turn of the century. A means was seen to accomplish these ends by combining the old idea of psychoneural interaction with the concept of mental phenomena as emergent properties delineated in operational terms (Sperry, 1952). Especially in the area of mind-brain relations, one tends to discover that nothing is ever new. As yet, however, I have not been able to find any proposals that attempt to resolve the problem of subjective experience in quite the same way—prior, that is, to 1965.

It is generally accepted that emergent phenomena elsewhere in the universe and at all levels of organization possess causal potency in the objective monistic realm of science. One may ask accordingly why the cerebral emergents postulated to be involved in conscious awareness have been denied similar status and have been set off as exceptions to the rule in the currently prevailing parallelistic and epiphenomenal interpretations. To concur with the idea that consciousness is probably an emergent phenomenon on the one hand, as Bindra appears to do, and then to object to giving the emergent properties a causal role in this special case leads to a kind of dualistic position that requires some explanation. The present interpretation avoids this dualistic implication with the insistence that conscious phenomena are no exception in this respect, and like other emergent properties possess causal potency in material events. Only in this somewhat indirect and inferential sense was my first proposition that conscious awareness is an emergent phenomenon taken to imply support for the second proposition that consciousness has a causal influence.

The objection that the hypothesis remains vague on details is of course valid and must probably continue to apply to any proposal regarding mind-brain relations for some time to come. The proposed scheme properly interpreted, however, ought to appear somewhat less vague than alternative theories of mind available to date. Perhaps it may help to clarify the present scheme if we spell out more precisely how it compares with and how it differs from some of our current alternatives (Boring, 1942; Hook, 1961b; Kohler & Held, 1949). One common view, of long standing, conceives consciousness as an epiphenomenon, that is, something that proceeds separately but in parallel with the brain process, perhaps caused or generated by it, but never acting back upon it. The degree and quality of separation between the epiphenomenal subjective experience on the one hand and the material cerebral activity on the other may vary in the minds of different proponents ranging from a direct passive by-product to parallelistic phenomena in a separate mental realm. As an epiphenomenon, consciousness may or may not be related to

the neural activity as an emergent property, but there has been general agreement that subjective epiphenomena do not act back upon, or in any way influence, the brain's physical activity. The present view contends, in contradiction, that the conscious phenomena as direct emergent properties of the cerebral activity do interact with and causally determine the brain process.

In older views that lost favor with the advent of behaviorism and dialectic materialism in the early 1900s (Boring, 1942), the mind was believed to interact with and control the physical brain process. These older forms of mental interactionism differed from the present in that mind and consciousness were not conceived as holistic properties of the material process but more separately in more metaphysical and dualistic terms. I have described earlier the present view (Sperry, 1965, 1966b) as a return toward a form of mentalism. However, it seems as accurate and perhaps preferable for reasons outlined below to describe the present concept as a compromise position intermediate between dualistic mentalism and the classical monistic materialism that excludes mental forces.

Another interpretation of mind that has a large following has described consciousness as an inner view or inner aspect of the brain process. In this doctrine there exists only the one material brain process which may be viewed either objectively from the outside or subjectively from within. According to the inner aspect hypothesis, a complete causal account of the brain process is possible in objective terms without reference to the subjective aspects. The present view differs in contending that a complete objective explanation must include the conscious phenomena which are an integral part of, and have causal influence in, the material process, rather than being just a different aspect of it. It contends further that a complete objective description that included the spatial and temporal pattern features of the neural activity, as such, would perforce include the mental properties.

In the related "identity" theory of mind, subjective phenomena have been postulated to be identical with the neural activity, something that is supposed to become apparent if one looks at and talks about the mind-brain relation in the appropriate way. The present view differs in postulating that the phenomena of subjective experience as pattern properties are distinct, different from, and more than the neural activities of which they are composed. It may be noted further that the same conscious effect in the present view can be produced by different neural events on different occasions in different neural contexts provided the critical operational result at the holistic functional level is the same. Similarly, the identical neural process could have different conscious properties on different occasions depending on the contextual activity in which it appears. The identity theory might conceivably be expanded to include the kind of psychoneural interaction we now envisage, but not without significant changes in its present approach and formulation.

The Gestalt school of psychology was developed around the application of holistic principles in behavior with special focus on perception. In respect to mind-brain relations, the concern was to find brain correlates that would simulate or correspond with the observed configural effects in subjective experience. Isomorphic electric field patterns were postulated, based on hypothesized volume electric conduction effects in the cerebral cortex (Kohler & Held, 1949). The theory remained parallelistic and epiphenomenal; there was no contention as in the present scheme that the subjective phenomena themselves acted back upon the cerebral physiology. The present view also differs in that it does not require an isomorphic type of relationship between the cerebral process and its subjective properties. Other kinds of relations are conceivable for producing subjective meaning (Sperry, 1952, 1966a, 1967); conscious meaning can be conceived in functional operational terms, rather than in terms of isomorphic correspondence.

One hears the argument that mental phenomena could not possibly interact with and control the concomitant neural activity without violating thereby some of the laws of physics. In the proposed scheme there is no more a violation of physical laws than when the constituent elements of any object or system are controlled by the gross properties of the system as a whole. This type of control is universal in nature and simply requires that the laws of physics be recognized to include the fact that the position in time and space of objects and their parts is as critical in the determination of causation as are the intrinsic properties.

A related argument contends that the mental events cannot function as links in the causal chain of brain activity because of the timing wherein the mental experiences and their neurophysiological correlates occur in perfect simultaneity. Thus, on temporal grounds as well, it

is inferred that mental phenomena could not intervene in the chain of neural events and that the latter must accordingly be determined solely by physiological forces. This argument, like the foregoing, fails to take into account the type of causal control that prevails in the holistic form of psychophysical interaction wherein the emergent properties of the whole determine the fate of the parts. This kind of control is not ruled out by simultaneity in the mental and the neural process.

Another current theory with numerous adherents, not unrelated to the foregoing, has maintained that the problem of mind-brain relations is only an artifact of semantics, a pseudoproblem that arises out of the language we use in thinking about the question, and one that dissolves provided one approaches and states the problem in corrected terminology. The proposed interpretation, by contrast, recognizes the reality of both the cerebral events and the conscious mental phenomena, as well as the critical importance of trying to explain and understand their relationship.

In brief, the present concept differs from other current theories as follows: Instead of viewing consciousness as something metaphysical or epiphenomenal, or as an inner aspect of the cerebral process, or as being identical or parallel with it, or as just a pseudoproblem of semantics—as has variously been the custom in behavioral science for over half a century—it is now conceived as an integral dynamic property of brain activity having causal control influence on the flow pattern of neural excitation. Whereas all interpretations of conscious mind that heretofore have received any serious acceptance in brain research would permit the scientist to ignore consciousness in his objective models of cerebral function, the present view gives consciousness a functional role and requires that the subjective phenomena be included in the causal sequence for a complete explanation. To repeat, this represents a return toward mentalism except that the mental forces are not viewed in any metaphysical, preternatural, nonmaterial, epiphenomenal, or other dualistic sense.

In the proposed scheme, the mental phenomena and the physical brain process are seen to be mutually interdependent. Neither is primary nor ultimate to the exclusion of the other. In this sense, as well as in others mentioned above and earlier (Sperry, 1969), the present scheme represents a compromise between the older philosophies of mentalism and materialism (Boring, 1942; Hook, 1961b). I can only disagree with Bindra regarding his pronouncement that such a compromise is impossible—even in the form in which he reinterprets it to imply degrees of dualism. Dualistic mentalism has been expressed in different forms with different implications by different schools and the same can be said for materialism. It is not a simple all-this or all-that alternative. Within the context of the present scheme, it may be seen that as systems get more complex, the phenomena and properties of the whole system tend to become increasingly remote in quality from those of its elementary parts. A wide spectrum of qualitative differences is evident that approaches an extreme in the case of the brain and could be taken as degrees of dualism.

The present hypothesis constitutes a compromise in the following sense as well: Proponents of materialism may claim that inasmuch as the mental phenomena are conceived to be properties of the material brain process, the theory therefore falls within the scope of materialist doctrine. Proponents of mentalism on the other hand can equally well claim that if the subjective phenomena of inner experience are recognized in this scheme to be distinct mental phenomena in their own right with governing influence on the sustaining brain process, then this is all they would demand as mentalists.

The kind of control that the mental properties are conceived to exert over neural activity, described as emergent or holistic psychophysical interaction, seems not to have been correctly comprehended by Bindra. It is not merely a controlling action of one brain center upon another as he interprets it, although this type of control undoubtedly exists and may be involved in many or most conscious brain processes. This, however, is not the point. In an earlier explanation directed at a nontechnical audience (Sperry, 1965), I tried to spell out the concept in the simplest possible terms by reference to the command hierarchy of causal controls that exists within the brain, that is, the forces within forces that extend from the level of the elementary particles of brain nucleons and electrons on up successively through atoms, molecules, neurons, nerve circuits, and cerebral processes in action. The course of events within each subsystem, relative to the rest of the brain, is governed by the properties of the higher level systems within which each is embedded. This same control principle is held to apply also with respect to conscious phenomena and the neural events out of which they are built at the top of the hierarchy.

I think of no simpler example to illustrate this form of control than the one of the wheel rolling downhill in which the displacement in time and space and the subsequent fate of the entire population of atoms, molecules, and other components within the system are determined very largely by the holistic properties of the whole wheel as a unit, like its shape, size, weight, etc. Similarly, the cerebral process for a unitary mental experience, like a visual image, is presumed to possess an inertia, coherence, and related dynamic properties as a unit that cause it to behave and to be treated in cerebral dynamics as a distinct entity. As in the wheel, the component parts of such an excitatory neural process are carried along and thus controlled by the dynamic properties of the whole system. That is, the process as a whole is facilitated, suppressed, perseverated, or whatever, on the basis of its holistic qualities. The present interpretation suggests that visual images, for example, would be perceived, stored, and retrieved primarily in terms of their subjective properties. It follows directly that the cerebral machinery must be so organized that it can use the subjective qualities of the excitation process.

Holistic interaction is presumed to apply in cerebral activity whether the neural events in question are of the conscious variety or not. The holistic form of control is ubiquitous, applying at all levels of neural integration, but it is presumably only in higher order cerebral activity that the conscious properties emerge. It does not follow, however, as inferred by Bindra, that it is correct to identify higher order neural processes with consciousness. Many high order neural functions, as in the cerebellum, for example, as well as in the cerebrum, presumably lack conscious quality. It is not mere complexity or high order organization that in this scheme endows a neural event with conscious awareness. It is rather the specific operational design of the cerebral mechanism for the particular conscious function involved. The neural mechanisms for conscious experience are not just more complex, they are specifically structured on an operational, functional basis to create particular sensations, percepts, and feelings, and to provide a rapid representation of external reality. This point also had been brought out earlier (Sperry, 1952, 1964, 1966a) and probably was not sufficiently reviewed in the rather condensed 1969 statement.

Misinterpretation of the nature of holistic control in the discussion of Bindra (1970) entangled with the further misinterpretation of "conscious" activity by equating it with "higher order" phenomena understandably makes for some puzzlement. The subsequent circumlocution based on these misconceptions, by which it is then inferred that the proposed hypothesis is circular and tautological, does not help to clear the confusion. Correctly grasped, the proposed concept of consciousness is simple and straightforward with no circular reasoning. In fact, an encouraging feature of the scheme is that, once seen, it appears almost obvious, eliciting responses such as "Of course!" "What else?"

Bindra is quite correct in stating that I have not defined in concrete terms the exact organizational features of the neural process responsible for the conscious effects, nor advanced a definitive proof that subjective experience influences neural actions. Anyone who does so will, of course, no longer have a hypothesis but the proven answer. Most of us who work in the area have become reconciled to seeking a solution to the problem of consciousness only through a series of successive approximations rather than by any single breakthrough that is complete and definitive. However, the request for an explanation in more concrete terms was provided to some extent in my 1965 account where the sensation of pain from a phantom limb was used to illustrate the proposed interpretation as it applied to the simplest kind of mental event. The same reasoning holds for perception and for more complex conscious functions. As another example, consider the perception of a melody whistled or tapped on a piano. At the apparent location of the experienced tones one finds only a succession of mechanical events producing air turbulence and pressure waves. The musical notes of subjective experience are part of the brain activity itself and are located within the brain. Also within the brain there will be, of course, the constituent phenomena of neuroglial excitation of which the auditory sensations are composed, plus also in turn the constituent molecular and atomic phenomena, etc. The overall pattern of this excitation hierarchy is presumed to possess, in proper context, the experienced subjective properties with their causal effects that enable such an excitatory pattern to be treated in cerebral function as a unit. The subjective experience is conceived to be but one among a large variety of holistic properties with causal effects in the brain process existing and operating at different levels of organization.

The presence and importance of such configurational, pattern, or gestalt properties in

the control of cerebral activity at the perceptual and cognitive level seems amply supported on numerous grounds. The comment of the old maestro is relevant: "Anybody can play the notes; it's the shaping of those intervals of silence between the notes that makes the difference." Gestalt principles apply as well to the organization of conscious elements into higher functions above the level with which we are here concerned. The interrelated, relativistic, organizational features of cerebral excitation seem primary and essential from the subjective standpoint. Exactly how these properties of the brain mechanism operate in terms of the neural circuitry remains to be determined. Like alternative theories of consciousness, the present suggests a solution in broad principle only. To determine exactly the precise features of circuit design by which cerebral activity produces its conscious effects remains, as was stated, a central challenge for the future. One merely hopes at this stage that future experimental and intellectual efforts, if focused within the guidelines of the present hypothesis, will be more effective than if directed on the assumption that mind is just an epiphenomenon, an inner aspect, a pseudo-problem, etc.

On paper at least the hypothesis appears to have a number of distinct attractions: By giving consciousness a functional role in brain activity with causal control potency, we would restore mind, including subjective mental phenomena of all kinds, to the brain of objective science. This in turn would largely resolve the long-standing dichotomy that has existed within psychology between the objective physiological approaches to behavior and those based on subjective phenomenology with concern for inner feelings and experience. The scheme also offers a compromise, as described, between the philosophies of materialism and mentalism, proposing a mutual mind-matter interaction in the upper realms of a single continuous hierarchy. It would provide further a partial answer to the old issue of free will versus determinism (Hook, 1961a). By bringing subjective phenomena into the causal sequence of the cerebral decision-making machinery, additional degrees and kinds of freedom are introduced that are not evident where the brain process is conceived merely in terms of electrochemical and neurophysiological determinants. Determinism still holds in the sense that all decisions are caused, but the kind of causation involved becomes much more palatable to proponents of indeterminism.

Bringing subjective phenomena into the realm of science, and vice versa, also would help greatly to bridge the traditional chasm between science and the humanities. Subjective values, for example, and value judgments become objective causal factors in the global chain of control, their origins, structure, and function no longer out of bounds to science on philosophical or other grounds. No loss, aesthetic or spiritual, need be incurred in this approach provided the fallacies of reductionism are avoided (Sperry, 1965). All in all, the proposed interpretation with its unifying concept of the mind/brain and mind/reality interfaces provides a basis for a comprehensive view of human nature and of man's place in the total scheme with far-reaching ideological implications. Regardless of whether this particular interpretation proves right or wrong, the mind-brain sciences are gradually closing in on some of these age-old problems, the answers to which now assume special relevance in today's renewed search for meaning.

REFERENCES

BINDRA, D. The problem of subjective experience: Puzzlement on reading R. W. Sperry's "A modified concept of consciousness." *Psychological Review*, 1970, **77**, 581–584.

BORING, E. G. *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century-Crofts, 1942.

HOOK, S. (Ed.) *Determinism and freedom*. New York: Collier Books, 1961. (a)

HOOK, S. (Ed.) *Dimensions of mind*. New York: Collier Books, 1961. (b)

KOHLER, W., & HELD, R. The cortical correlate of pattern vision. *Science*, 1949, **110**, 414–419.

SPERRY, R. W. Neurology and the mind-brain problem. *American Scientist*, 1952, **40**, 291–312.

SPERRY, R. W. Problems outstanding in the evolution of brain function. In, *James Arthur lecture*. New York: American Museum of Natural History, 1964.

SPERRY, R. W. Mind, brain and humanist values. In J. R. Platt (Ed.), *New views of the nature of man*. Chicago: University of Chicago Press, 1965.

SPERRY, R. W. Brain bisection and mechanisms of consciousness. In J. C. Eccles (Ed.), *Brain and conscious experience*. New York: Springer-Verlag, 1966. (a)

SPERRY, R. W. Mind, brain and humanist values. (Condensed version.) *Bulletin of the Atomic Scientists*, 1966, **22**, 2–6. (b)

SPERRY, R. W. Split-brain approach to learning problems. In G. C. Melnechuk & F. O. Schmitt (Eds.), *The neurosciences: A study program*. New York: Rockefeller University Press, 1967.

SPERRY, R. W. A modified concept of consciousness. *Psychological Review*, 1969, **76**, 532–536.

(Received August 4, 1970)